

KAYTUS

KR6288V2

Hyperscale Training Platform



Overview

KR6288V2, the 6U hyperscale training platform equipped with dual 4th Gen Intel Xeon Scalable Processors or AMD EPYC™ 9004 Series Processors and 8x fully connecting GPUs, features industry-leading performance, ultimate I/O expansion, and ultrahigh energy efficiency. The precisely optimized system architecture with 4x CPU to GPU bandwidth, up to 4.0Tbps networking bandwidth, 8TB system memory, and 300TB massive local storage can fully satisfy the communication and capacity demands of multi-dimensional parallelism training for giant-scale models. 12 PCIe expansions can be flexibly configured with CX7, OCP3.0, and multiple SmartNICs, making it an ideal solution for both on-premises and cloud deployment. It is built to handle the most demanding AI computing tasks like trillion-parameter Transformer model training, massive recommender systems, AIGC, and Metaverse workloads.

Benefits

■ Unprecedented Performance

- > Powered by 8x fully connecting GPUs in a 6U chassis, TDP up to 700W
- > Support 2x 4th Gen Intel® Xeon® Scalable Processors or AMD EPYC™ 9004 Series Processors
- > Industry-leading performance with high AI performance by 3 times enhancement

■ Leading Architecture Design

- > Lightning-fast intra-node connectivity with 4x CPU to GPU bandwidth improvement
- > Ultra-high scalable inter-node networking with up to 4.0Tbps non-blocking bandwidth
- > Cluster-level optimized architecture, GPU: Compute Network: Storage Network = 8:8:2

■ Optimized Energy Efficiency

- > Extremely low air-cooled heat dissipation overhead, less fan, higher power efficiency
- > 54V, 12V separated power supply with N+N redundancy reducing power conversion loss
- > Direct liquid cooling design with more than 80% cold plate coverage, PUE≤1.15

■ Multi-scenarios Adaptation

- > Full modular design and extremely flexible configurations satisfying both on-premises and cloud deployment
- > Easily harness large-scale model training, such as GPT-3, LLaMA and Stable diffusion.
- > Diversified SuperPod solutions accelerating the most cutting-edge innovation including AIGC, AI4Science and Metaverse

Product Specifications

Item	KR6288-X2-A0-R0-00	KR6288-E2-A0-R0-00
Height	6U	
GPU	8x GPUs, TDP up to 700W per GPU	
Processor	2x 4 th Gen Intel® Xeon® Scalable Processors, TDP 350W	2x AMD EPYC™ 9004 Series Processors, Max cTDP 400W
Memory	32x DDR5 DIMMs, up to 4800MT/s	24x DDR5 DIMMs, up to 4800MT/s
Storage	24x 2.5' SSD, up to 16x NVMe U.2	
M.2	2x Onboard NVMe/SATA M.2 (optional)	2x Onboard NVMe M.2 (optional)
PCIe Slot	Support 10x PCIe Gen5 x16 slots. One PCIe Gen5 x16 slot can be replaced with two x16 slots (PCIe Gen5 x8 rate). Optional support Bluefield-3, CX7, and various SmartNICs	
RAID	Optional support RAID 0/1/10/5/50/6/60, etc., support Cache super capacitor protection	
Front I/O	1x USB 3.0, 1x USB 2.0, 1x VGA	
Rear I/O	2x USB 3.0, 1x MicroUSB, 1x VGA, 1x RJ45	
OCPI	Optional support 1x OCP 3.0, support NCSI	
Management	DC-SCM BMC management module with Aspeed 2600	
TPM	TPM 2.0	
Fan	GPU region: 6x 54V hot-swap fans with N+1 redundancy CPU region: 6x 12V hot-swap fans with N+1 redundancy	
Power	2x 12V 3200W and 6x 54V 2700W, Titanium CRPS PSU with N+N redundancy	
Size	Width: 447mm, Height: 263mm, Length: 860mm	
Weight	Net weight 92kg(Cross weight: 107kg)	
Environmental Parameters	Working temperature: 10°C~35°C; Storage temperature: -40°C~70°C Working humidity: 10%~80% R.H.; Storage humidity: 10%~93% R.H.	